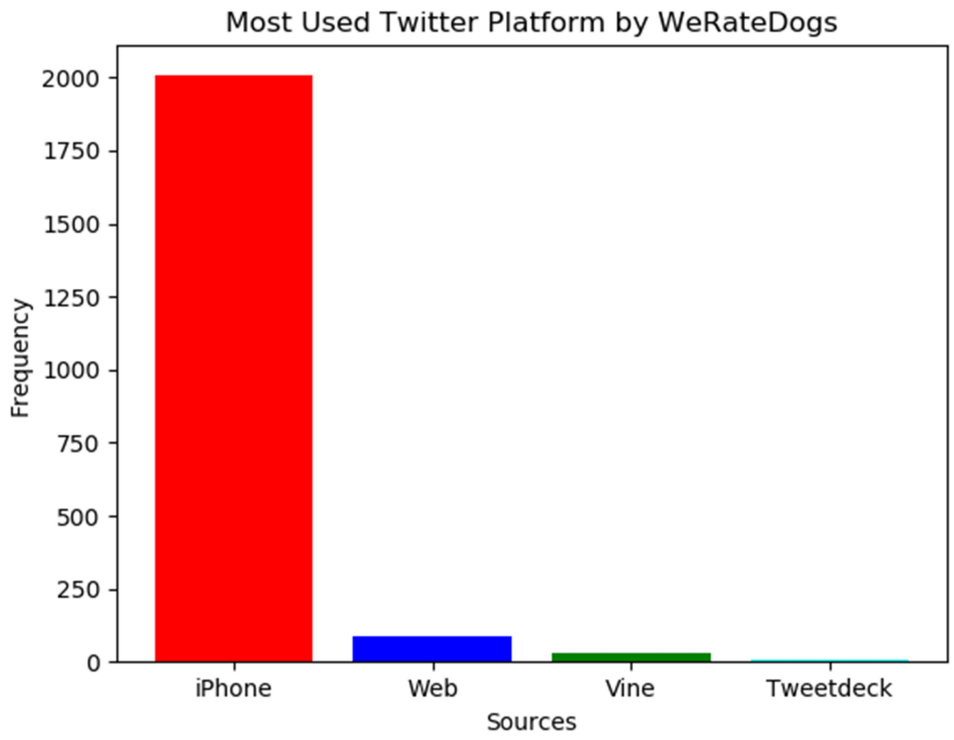
After gathering and cleaning the data, I looked for potential information that we can gather. Following are the observations or insights that I got.

1. Most Used Twitter Platform by WeRateDogs

   I was curious about the client which WeRateDogs used for tweeting. For this I used the 'source' & counted the unique values and how many times these occur in the dataset.

   Following chart was made on the data acquired:



   From the charts it is evident that most of the tweets have been made from **Twitter for iPhone**.

2. Which dog types are most famous?

   For this I used 'group_by' on the 'dog_type' column of data and had the following results about the average values.

Index	rating_numerator	favorite_count	retweet_count
None	10.5601	8129.73	2446.36
doggo	11.7701	17456.6	6040.83
floofer	11.8889	10516.8	3878.56
pupper	10.7354	6907.48	2311.82
puppo	12.0417	20746.5	6066.25

From this we can infer that tweets with no dog types are least famous and least liked and retweeted by users whereas puppo dogs receive the highest average rating and favourite as well as retweet count. So having a puppo photo surely increases your chances of getting famous.

3. Top 5 Dog Breeds which are most favorites.

   Using the same method as above on 'dog_breed' column and sorting according to 'favorite_count' gives the following result.

Index	rating_numerator	favorite_count	retweet_count
soft-coated_wheaten_terrier	14	45074	11327
Rhodesian_ridgeback	13	41010	8370
miniature_pinscher	12	39210	9075
Welsh_springer_spaniel	13	37694	8213
black-and-tan_coonhound	13	32790	10096

4. Most Common Dog Name?

None	753
Lucy	11
Charlie	11
Cooper	10
Oliver	10
Penny	9

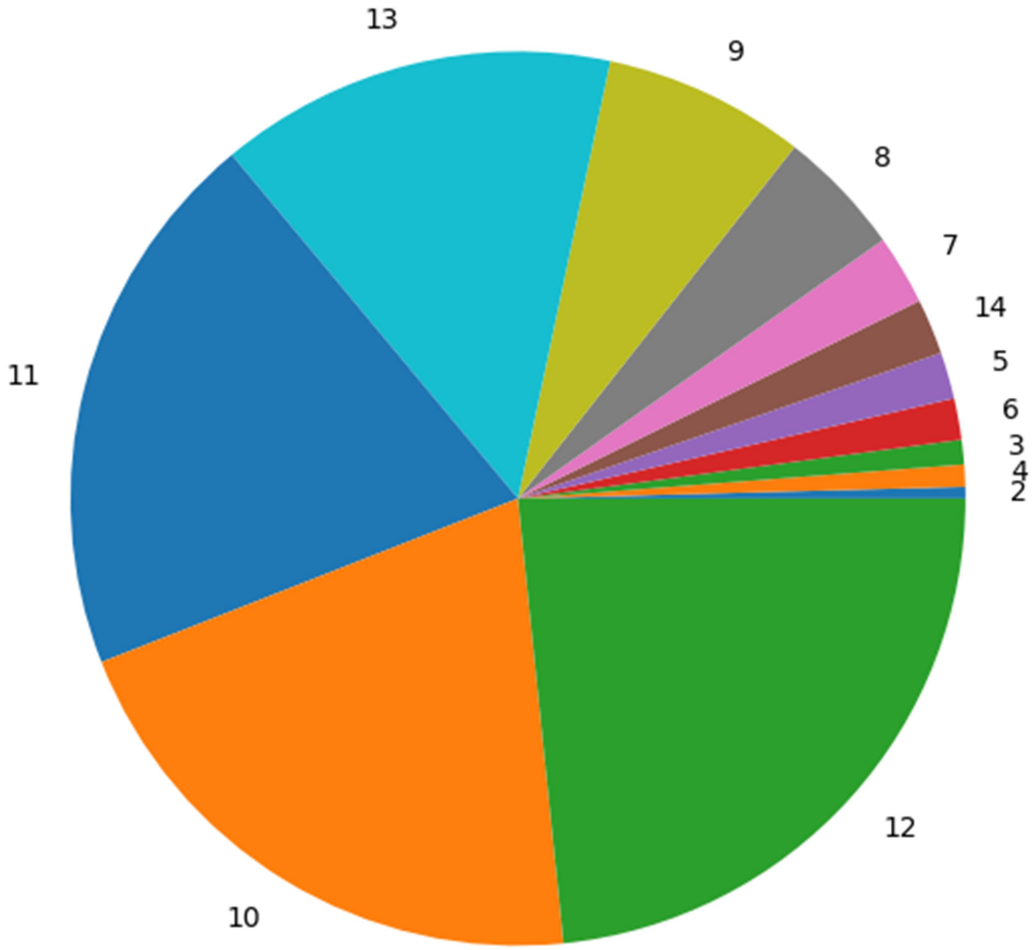
Most of the tweets did not contain a name. Out of the tweets that did top 5 repeated names are Lucy, Charlie, Cooper, Oliver & Penny.

5. Average Likes, Retweets & Ratings of tweets.

avg_likes	float	1	8533.26507713885
avg_rating	float	1	10.649836372136512
avg_retweets	float	1	2625.1706404862084

We can infer from this data that if our dogs photo is tweeted by WeRateDogs , It will always reach a larger audience.

6. Distribution of Ratings



We can see that expected ratings of pictures can be in the range of 10-13 and higher ratings are given to very less dogs as well as less ratings.

Conclusion:

Since we only worked with textual data, information was limited. By applying machine learning models to these dataset we can even predict how much retweets or favorites counts that a photo can get. Also extracting data from text can be improved which could have made the data more refined.